# Genome-wide association study identifies 74 loci associated with educational attainment

Okbay, Aysu; Beauchamp, Jonathan P; Fontana, Mark Alan; Lee, James J; Pers, Tune H; Rietveld, Cornelius A

*Published in:*
Nature

*DOI:*
10.1038/nature17671

*Publication date:*
2016

*Document Version*
Peer reviewed version

Author Accepted Manuscript version of Okbay, Aysu et al. "Genome-wide association study identifies 74 loci associated with educational attainment". Nature. 2016, 533(7604). 539-542. https://doi.org/10.1038/nature17671

# Title: Education-associated SNPs are enriched for brain function and disorders

**Authors:** All authors and their affiliations appear at the end of the paper*

*Correspondence to: Daniel Benjamin (daniel.benjamin@gmail.com), David Cesarini (dac12@nyu.edu), Philipp Koellinger (koellinger@ese.eur.nl), or Peter Visscher (peter.visscher@uq.edu.au).

**One Sentence Summary:** A GWAS meta-analysis of almost 300,000 individuals identifies 74 independent SNPs associated with educational attainment, pointing to genes, pathways, and tissue types known to be involved in brain development and structure, and the architecture of axon-dendrite connections.

**Summary:** Educational attainment has important consequences for a range of economic, social, and health outcomes. Although it is known to be substantially heritable in developed countries, few specific variants have been credibly identified. Here we report a large genome-wide association study of 293,723 individuals. We identify 74 independent genome-wide-significant SNPs. Genetic associations are enriched for genes, pathways, and tissue types known to be involved in brain development and the architecture of axon-dendrite connections. The genes in our identified loci are preferentially expressed in neural tissue, particularly during prenatal development, and are implicated in brain size and structure. The genetic loci disproportionately reside in genomic regions regulating gene expression in the fetal brain. A genome-wide pattern of education-increasing variants is associated with increased cognitive performance, intracranial volume, and height, an increased risk of schizophrenia and bipolar disorder, decreased risk of Alzheimer's, and lower neuroticism.

## Main Text:

Educational attainment (EA) has important consequences for a range of economic, social, and health outcomes[1,2]. Discovery of genetic variants associated with EA may yield insights into the biological pathways underlying phenotypes[3], such as neuropsychiatric conditions[4] and brain morphology, that may exhibit genetic overlap with EA. In addition, a "polygenic score" for EA—an index that aggregates the effects of many genetic variants—can be useful in empirical research as a control for unobserved heterogeneity, or to study gene-environment interactions[5].

Across developed countries, genetic factors as a whole account for a modest to moderate share of the variation across individuals, roughly 40% on average[6]. Yet as with other behavioral traits, few credible genetic associations have been identified. The largest genome-wide association study (GWAS) meta-analysis of EA to date had a discovery sample of $N$ = 101,069 and identified 3 genome-wide significant single nucleotide polymorphisms (SNPs)[7]. Here, we

report the results of a GWAS meta-analysis in a sample of $N$ = 293,723, which brings the number of SNPs to 74. Our sample size enables us to obtain informative results from many additional analyses, such as providing evidence for a number of key biological pathways and quantifying genetic overlap with neuropsychiatric disorders and other phenotypes.

*Identification of 74 Lead SNPs*

We examined two phenotypes: a variable measuring the number of years of schooling completed (*EduYears*, $N$ = 293,723, mean = 14.33, SD = 3.61) and an indicator variable for college completion (*College*, $N$ = 280,007, mean = 0.29). We harmonized the study-specific measures of EA using the International Standard Classification of Education (ISCED 1997) scale. All analyses were performed at the cohort level using the 1000 Genomes Project reference panel[8] according to a public analysis plan. The estimation sample is restricted to individuals of European descent whose EA was assessed at age 30 or older. A uniform set of quality-control (QC) procedures were used, including but not limited to the best practices previously recommended[9]. *EduYears* was our primary outcome variable and was used for all follow-up analyses (such as biological annotation). The analyses are based on ≈9.3M SNPs that passed QC filters. We applied single genomic control at the cohort level and meta-analyzed results using sample-size weighting.

To define the number of approximately independent genome-wide-significant signals—henceforth "lead SNPs"—we first clumped the genome-wide-significant signals in PLINK[10] using an $r^2$ threshold of 0.1 and a distance threshold of 500 kb. We then further pruned the resulting set of index SNPs by checking for long range LD ($r^2 > 0.5$) between each pair on the same chromosome. The Manhattan plot for *EduYears* is shown in Fig. 1. The 74 lead SNPs, each representing a locus (which include the 3 previously identified SNPs[7]), are highlighted. Extended Data Fig. 1 provides Manhattan plots for the sex-stratified analyses of *EduYears* and the pooled *College* analysis. As in the earlier GWAS of EA[7] and other large GWAS of polygenic traits[11–13], the Q-Q plot of the meta-analysis (Extended Data Fig. 2) exhibits inflation ($\lambda_{GC}$ = 1.28), consistent with a polygenic architecture. Extended Data Fig. 3 shows the distribution of the estimated effect sizes on *EduYears* of the lead SNPs. They are in the range 0.05 to 0.17 years of schooling (approximately 3 to 9 weeks) per allele, with $R^2$ in the range 0.01% to 0.035%.

Given the very small $R^2$'s, an important concern is that results could be driven by very subtle population stratification that remains even after the stringent controls used by the cohorts (see Supplementary Information section 1.4)[14]. As a standard first test, we calculated how often the signs of the regression coefficients aligned across two independent meta-analyses: a within-family analysis conducted in dizygotic (DZ) twin pairs from four cohorts (STR, MCTFR, NTR, and QIMR; $N$ = 8,233 DZ twin pairs in total), and the GWAS meta-analysis excluding these cohorts. Of the 66 lead SNPs available in all four cohorts, the signs aligned 62% of the time, significantly more than the 50% expected under the null that the results are entirely due to stratification ($P$ = 0.032). Additionally, we quantified the amount of stratification in the GWAS estimates using LD Score regression[15]. The results indicate that 8% of the observed inflation in the mean $\chi^2$ is accounted for by bias rather than polygenic signal (Extended Data Fig. 4a), suggesting that stratification is not a major concern in the analyses that follow.

*Genetic Overlap with Other Phenotypes*

We used bivariate LD Score regressions[15] to estimate the genetic correlation between *EduYears* and each of 14 phenotypes for which we obtained published GWAS results: 3 neuropsychiatric conditions (bipolar disorder, schizophrenia, Alzheimer's), 2 anthropometric traits (BMI, height), 2 behavioral traits (neuroticism, cognitive performance), intracranial volume (ICV), and volumes of 6 subcortical brain structures adjusted for ICV. As shown in Fig. 2a (and Supplementary Table 3.1), the genetic correlations are statistically distinguishable from zero for all the non-subcortical-structure phenotypes. Overall, having more EA-increasing alleles is associated with increased cognitive performance, intracranial volume, and height, an increased risk of schizophrenia and bipolar disorder, and a decreased Alzheimer's risk and lower neuroticism. Bulik-Sullivan et al.[4] used the earlier EA GWAS[7] results to study some of the same phenotypes. For the overlapping phenotypes, our estimates are similar to theirs. However, we observe the largest genetic correlations for three phenotypes not studied by Bulik-Sullivan et al.: cognitive performance ($\hat{\rho}_G$ = 0.75), neuroticism (−0.41), and intracranial volume (0.34).

Next, we looked up our lead SNPs in the phenotypes' GWAS results (Fig. 2b and Extended Data Fig. 5). When we test the 74 SNPs jointly for association, we reject the null hypothesis of no enrichment at $p < .05$ for 10 of the 14 phenotypes (all the exceptions are subcortical brain structures). We typically find stronger enrichment for phenotypes whose estimated genetic correlation is further from zero. An intriguing exception is schizophrenia: it is very strongly enriched for our lead SNPs, but its genetic correlation is close to zero because the SNPs have opposing affects on the phenotype: roughly half with concordant effects on *EduYears* and half discordant. In these lookup exercises, enrichment could arise for several reasons, including (i) pleiotropy, (ii) co-localization of distinct causal genes (e.g., the genes associated with education might be near the genes associated with schizophrenia), or (iii) the phenotype mediates genetic effects on *EduYears* (this last being most plausible for cognitive performance and neuroticism).

We examined whether individual SNPs' *P*-values reached a significance threshold of 0.05/74 across the 14 phenotypes. We observed a total of 25 significant associations (Supplementary Table 3.4). Several of these replicate previously known associations with BMI[11], cognitive performance[16], height[13], intracranial volume[17,18], and schizophrenia[12]. Others are not within 1,000 kB of any genome-wide significant SNPs reported in previous GWAS; for example, we observe one such association each with Alzheimer's, cognitive performance, neuroticism, schizophrenia, hippocampal volume, and intracranial volume (Extended Data Fig. 6, Supplementary Table 3.3).

*Biological Annotation*

To identify candidate biological pathways and genes, we probed the loci identified in our analyses with a number of bioinformatic tools.

Using an empirical Bayesian hierarchical model applied to all SNPs regardless of statistical significance[19], we estimated that DNase I hypersensitive regions (i.e., regions related to gene expression) in the fetal brain are enriched for *EduYears*-associated SNPs by a factor of ≈5 (Extended Data Fig. 7a,b). As a robustness check, we reapplied our method using an alternative set of annotations from the Roadmap Epigenomics data and found quantitatively similar results (Supplementary Table 4.2.3). We also found that *EduYears*-associated SNPs are enriched positively in regions that are evolutionarily conserved in mammals and negatively in regions

that are transcriptionally repressed (Extended Data Fig. 7a and Supplementary Table 4.4.1), results that accord with a trend observed in the analysis of other phenotypes[19,20].

Our two tissue-level analyses also implicate brain-specific gene regulation. First, we used LD Score regression to estimate an enrichment factor—the fraction of the *EduYears* genetic variance explained by SNPs residing in regions associated with histones marked specifically in a particular type of tissue, divided by the fraction of SNPs bearing this tissue-specific annotation—for each of 10 tissue types (Supplementary Table 4.4.2)[20]. We also applied the same procedure to three phenotypes—height[13], body mass index (BMI)[11], and waist-to-hip ratio adjusted for BMI (WHR)[21]—that have been studied using samples sizes comparable to ours. A striking pattern in Extended Data Fig. 8 is that each of the 10 tissue types in this analysis, with one exception, is more enriched by WHR and height than by *EduYears* and BMI. The exception is the central nervous system, which is most enriched by *EduYears* (~3-fold). This finding is consistent with Finucane et al.[20], which contains an application of the same method to the earlier EA GWAS[7] results. Second, turning to direct measurements of gene expression, we tested the 37 adult tissues from the Genotype-Tissue Expression Project[22] for enrichment by mRNA transcribed from genes within loci defined by the enrichment-analysis tool DEPICT[23]. The 13 tissues that are components of the central nervous system are the only ones that are significantly enriched (Extended Data Fig. 9a and Supplementary Table 4.5.2).

We examined the functions of genes that overlap DEPICT-defined loci, centering on SNPs attaining $P < 1\times10^{-5}$ (Supplementary Tables 4.1.5 and 4.5.1). We used the same tool to nominate candidate functional genes within our loci. DEPICT identified 283 sets of functionally related genes (FDR < 0.05) (Supplementary Table 4.5.1). To help interpret the results, we partitioned the gene sets into 34 clusters (shown in Fig. 3a) based on degree of gene overlap. The gene sets map onto many processes of neural development: the proliferation of neural progenitor cells and their specialization (*npBAF complex*), the migration of new neurons to the different layers of the cortex (*forebrain development*, *abnormal cerebral cortex morphology*), the projection of axons from neurons to their signaling targets (*axonogenesis*, *signaling by Robo receptor*), the sprouting of dendrites and their spines (*dendrite*, *dendritic spine organization*), and neuronal signaling and synaptic plasticity throughout the lifespan (*voltage-gated calcium channels*, *synapse part*, *synapse organization*). In light of the moderate genetic correlation between *EduYears* and intracranial volume (Fig. 2a), we note that the cluster named after *abnormal cerebral cortex morphology* contains sets of genes that lead to pathological phenotypes such as decreased brain size and irregular layering of the cerebral cortex when knocked out or otherwise perturbed in mice[24].

*Signaling by EGFR* is named after a pathway where mutations contribute to overproduction of glial cells[25], and one of its members emerged as the top gene set in a recent GWAS of intracranial volume[18]. Most of the clusters above and to the left of *signaling by EGFR* in Fig. 3a are clearly neural in nature; some of the clusters below and to the right are less obviously so. These latter clusters (*protein kinase binding*, *transcription cofactor activity*, *histone acetyltransferase complex*) may appear to represent routine housekeeping pathways of little specific relevance to cognition, but in fact recent studies point to the importance of transcriptional regulation and chromatin remodeling in polygenic neuropsychiatric disorders[26]. Indeed, many of our DEPICT-nominated genes that are involved in determining the size and structure of the cortex encode transcription factors and components of nucleosome remodelers (Supplementary Table 4.5.1).

As candidates for follow-up work, Table 1 and Supplementary Table 4.1 list a number of individual genes implicated by convergent lines of evidence from ten analyses (the table captions describe how the genes were selected). Two of the most strongly prioritized genes (*FOXO6*, *FOXP2*) encode proteins belonging to the forkhead family of transcription factors. *FOXP2* was the first gene to be implicated in language, and it has undergone two nonsynonymous substitutions since the divergence of the human lineage from other primates[27]. Several genes, including *MAPT*[28] and *HTT*[29], can mutate to cause neurodegenerative disorders. (Notably, our evidence does *not* implicate *APOE*/*TOMM40*, a gene linked to cognitive decline and Alzheimer's disease[30,31].) One of the first SNPs found to be associated with *EduYears* was rs9320913[7], and its nearest gene (*POU3F2*) was implicated in our subsequent study of cognitive performance[16]. This gene is the highest-ranking DEPICT-nominated member of several significantly enriched gene sets, including *decreased brain size*, *central nervous system neuron differentiation*, and *abnormal axon guidance*. Many of the genes are already known to be involved in axon genesis and guidance (*DCC*[32], *SEMA6D*[33], *PCDH17*[34].) A particularly interesting gene in this respect is *MAP4*, which is normally regarded as a non-neuronal microtubule-associated protein but nevertheless emerges as one of our top candidates and a predicted participant in the pathways *signaling by Robo receptor* and *axon guidance*.

Previous large-scale GWAS have observed a pattern: genes in which *de novo* mutations are known or believed to cause syndromic forms of a disease are enriched in GWAS-identified loci[12,13]. To test for this pattern in our results, we examined whether our 74 lead loci disproportionately overlap genes where *de novo* mutations can cause (a) intellectual disability (ID), a syndrome characterized by low levels of cognitive performance, (b) a large effect on liability to autism spectrum disorder (ASD), and (c) a large effect on liability to schizophrenia (SCZ). Fig. 3b shows that genes nominated by DEPICT related to *EduYears* exhibit substantial overlap with ID genes ($OR$ = 4.61), more so than DEPICT-nominated genes for the three anthropometric traits. Conversely, genes nominated by DEPICT for height and WHR exhibit substantial overlap with genes where mutations can cause syndromic skeletal disorders, much more so than *EduYears* genes. We also found that our *EduYears* loci harbor significantly more ID and ASD genes than randomly chosen sets of loci selected to match the *EduYears* loci in certain genomic properties[23]. Extended Data Fig. 9b shows the DEPICT-nominated *EduYears* genes where mutations have been implicated in ID, ASD, and SCZ.

Our results above implicate a critical role for brain-specific gene expression during prenatal development, but do not test for expression during prenatal versus other periods of life. We did so using the BrainScan Developmental Transcriptome[35] data. We found that the DEPICT-prioritized genes exhibit above-baseline expression in the brain throughout life but higher expression levels in the brain specifically during prenatal development (1.36-fold over postnatal, one-sided paired $t$-test $P = 4.2\times10^{-8}$; Fig. 3c).

*Polygenic prediction*

To assess the joint predictive power afforded by the GWAS results, we constructed polygenic scores for *EduYears* using sets of SNPs whose nominal $P$-values fall below a certain threshold, ranging from $5\times10^{-8}$ (only the genome-wide significant SNPs were included) to 1 (all measured SNPs were included). For weights, we re-estimated the SNP effects jointly in a multiple-SNPs model using GCTA-COJO (Supplementary Table 5.1; the all-SNPs score uses the meta-analysis weights). We constructed each version of the score twice, excluding either STR or HRS ($N$ = 9,339 and 8,538 unrelated individuals, respectively), and tested the predictive power of the scores in the excluded cohort (Fig. 4a and Supplementary Table 5.2). As expected for a

highly polygenic trait, the fraction of variance explained increases as the *P*-value threshold becomes less conservative. The score that includes all SNPs predicts 2.6% of the variance in the STR sample and 3.8% in the HRS sample. The sample-sized-weighted mean is 3.2%.

To assess the extent to which population structure in the replication samples might account for the explanatory power of the score, we used the method of Wood et al.[13] on the DZ-twin data in STR ($N$ = 3,515 pairs) to partition the variance of the score built with the SNPs meeting selected *P*-value thresholds into components due to real SNP effects from SNPs that are in LD, to real SNP effects from SNPs that are correlated due to population structure, and to estimation error (Supplementary Table 2.2). We find that the variance explained by real SNP effects is sizeable, both before and after residualizing the score on 10 PCs of the genotypic data (Fig. 4b,c). As an additional test, we regressed *EduYears* on the all-SNPs score using only the within-family variation in STR (Supplementary Table 2.2). The coefficient, 0.215 (s.e. = 0.081), is statistically distinguishable from zero ($P$ = 0.008), further indicating that the score captures real SNP effects.

While the explanatory power of the all-SNPs score seems smaller than expected (see Supplementary Information section 5.2), it is large enough to be useful for other studies. To illustrate, we study mediation of the effect of the all-SNPs score on EA in two of our cohorts. We find that cognitive performance can statistically account for 23-42% of the effect, and the personality trait "openness to experience" can statistically account for roughly 7% (Supplementary Tables 6.3 and 6.4, Supplementary Information section 6). We also used the all-SNPs score in STR (calculated excluding STR) to analyze how the strength of the association between the score and *EduYears* varies with birth cohort (Extended Data Fig. 10, Supplementary Information section 7). We find that the predictive power of the score for the oldest individuals in the sample (born in 1929) is almost twice as large as for the youngest (born in 1958) ($P$ = 0.004). This finding suggests that the expansion of educational opportunity in Sweden, similar to that which occurred in many developed countries during the 20th century, may have altered the complex causal pathways from genes to education in a way that produced greater equality in EA.

*Discussion*

This study of EA is the largest GWAS of a behavioral phenotype reported to date. Our results reveal strong parallels between EA and anthropometric and medical phenotypes such as height and schizophrenia. As with these other phenotypes, the GWAS methodology has become far more productive as sample sizes have increased[36]: for EA, from 3 genome-wide significant SNPs for $N \approx 100{,}000$[7] to 74 SNPs for $N \approx 300{,}000$. Large samples have enabled a range of statistically well-powered analyses, generating both confirmations of prior findings and many new discoveries. The advances we report here suggest that GWA studies of other behavioral phenotypes for which no credible associations have yet been established may have success as larger discovery samples become available.

Our study also allows us to further refine our understanding of how EA compares quantitatively to anthropometric and medical phenotypes along several dimensions. For instance, Rietveld et al.[7] found, as do we, that in terms of $R^2$, the explanatory power of the top *EduYears*-associated SNPs (in the range 0.02% to 0.03%) is roughly one-tenth as large as that of the top height- and BMI-associated SNPs (in the range 0.3% to 0.4%). As we expand the number of identified SNPs to 74, we find that outside the top 3 SNPs, the ratio of variance explained falls to a factor of roughly 6 for height and 2 for BMI.

Moreover, our biological analyses yield a range of novel and distinctive results when applied to EA. In particular, we identify biological pathways that relate EA to brain development (especially prenatally), neuropsychiatric disorders, and the architecture of axon-dendrite connections.

Our study begins to elucidate some of the complex pathways by which genetic variants influence EA. Such knowledge is valuable both for the identification of associations with more proximal phenotypes, for which large-sample GWAS are currently not feasible, and for enhancing our understanding of how environmental factors, including policies, interact with genetic effects.

**References and Notes:**


1. Cutler, D. M., Lleras-Muney, A. & Vogi, T. in *Oxford Handb. Heal. Econ.* (Glied, S. & Smith, P. C.) 124–163 (Oxford University Press, 2011).

2. Goldin, C. & Katz, L. *The Race between Education and Technology*. (Belknap Press, 2010).

3. Krapohl, E. *et al.* The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 15273–15278 (2014).

4. Bulik-Sullivan, B. *et al. An atlas of genetic correlations across human diseases and traits*. *bioRxiv* (2015). doi:10.1101/014498

5. Benjamin, D. J. *et al.* The promises and pitfalls of genoeconomics. *Annu. Rev. Econom.* **4,** 627–662 (2012).

6. Branigan, A. R., McCallum, K. J. & Freese, J. Variation in the heritability of educational attainment: An international meta-analysis. *Northwest. Univ. Inst. Policy Res. Work. Pap.* **13-09,** (2013).

7. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340,** 1467–1471 (2013).

8. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

9. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9,** 1192–212 (2014).

10. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

11. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

12. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

13. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186 (2014).

14. Rietveld, C. A. *et al.* Replicability and robustness of GWAS for behavioral traits. *Psychol. Sci.* **25,** 1975–1986 (2014).

15. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

16. Rietveld, C. A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 13790–13794 (2014).

17. Ikram, M. A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat. Genet.* **44,** 539–544 (2012).

18. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520,** 224–229 (2015).

19. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94,** 559–73 (2014).

20. Finucane, H. K. *et al. Partitioning heritability by functional category using GWAS summary statistics*. *bioRxiv* (Cold Spring Harbor Labs Journals, 2015).

21. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518,** 187–196 (2015).

22. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

23. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31,** 418–20 (2015).

24. Bogue, M. A., Grubb, S. C., Maddatu, T. P. & Bult, C. J. Mouse Phenome Database (MPD). *Nucleic Acids Res.* **35,** D643–9 (2007).

25. Read, R. D. Drosophila melanogaster as a model system for human brain cancers. *Glia* **59,** 1364–76 (2011).

26. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515,** 209–215 (2014).

27. Van der Lely, H. K. J. & Pinker, S. The biological basis of language: insight from developmental grammatical impairments. *Trends Cogn. Sci.* **18,** 586–595 (2014).

28. Goedert, M. & Spillantini, M. G. A century of Alzheimer's disease. *Science* **314,** 777–81 (2006).

29. Cattaneo, E., Zuccato, C. & Tartari, M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nat. Rev. Neurosci.* **6,** 919–30 (2005).

30. Davies, G. *et al.* Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Mol. Psychiatry* **20,** 183–192 (2015).

31. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 1977–81 (1993).

32. Stein, E., Zou, Y., Poo, M. & Tessier-Lavigne, M. Binding of DCC by netrin-1 to mediate axon guidance independent of adenosine A2B receptor activation. *Science* **291,** 1976–1982 (2001).

33. Dudanova, I. & Klein, R. Integration of guidance cues: parallel signaling and crosstalk. *Trends Neurosci.* **36,** 295–304 (2013).

34. Hayashi, S. *et al.* Protocadherin-17 mediates collective axon extension by recruiting actin regulator complexes to interaxonal contacts. *Dev. Cell* **30,** 673–87 (2014).

35. Brainspan. at <http://www.brainspan.org/>

36. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).

**Acknowledgements:**

This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC). The SSGAC seeks to facilitate studies that investigate the influence of genes on human behavior, well-being, and social-scientific outcomes using large genome-wide association study meta-analyses. The SSGAC also provides opportunities for replication and promotes the collection of accurately measured, harmonized phenotypes across cohorts. The SSGAC operates as a working group within the CHARGE consortium. This study was supported by funding from the Ragnar Söderberg Foundation (E9/11), the Swedish Research Council (421-2013-1061), The Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund of the Foundations of Human Behavior, and the NIA/NIH through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810, and T32-AG000186-23 to NBER. A full list of acknowledgments is provided in the supplementary materials. Data for our analyses come from many studies and organizations, some of which are subject to a MTA, and are listed in the supplementary online materials.

**Authors:**

Aysu Okbay[1,2,3,*], Jonathan P. Beauchamp[8,*], Mark A. Fontana[4,*], James J. Lee[5,*], Tune H. Pers[6,7,*], Cornelius A. Rietveld[1,2,3,*], Patrick Turley[8,*], Guo-Bo Chen[181], Valur Emilsson[9,10], S. Fleur W. Meddens[43,3,11], Sven Oskarsson[12], Joseph K. Pickrell[13], Kevin Thom[14], Pascal Timshel[15,16,7], Ronald de Vlaming[1,2,3], Abdel Abdellaoui[17], Tarunveer S. Ahluwalia[18,15,19], Jonas Bacelis[20], Clemens Baumbach[21,22], Gyda Bjornsdottir[93], Johannes H. Brandsma[23], Maria Pina Concas[24], Jaime Derringer[25], Nicholas A. Furlotte[26], Tessel E. Galesloot[27], Giorgia Girotto[28], Richa Gupta[29], Leanne M. Hall[30,31], Sarah E. Harris[32,33], Edith Hofer[34,35], Momoko Horikoshi[36,37], Jennifer E. Huffman[38], Kadri Kaasik[39], Ioanna P. Kalafati[40], Robert Karlsson[41], Augustine Kong[28], Jari Lahti[39,42], Sven J. van der Lee[2], Christiaan de Leeuw[43,44], Penelope A. Lind[45], Karl-Oskar Lindgren[12], Tian Liu[46], Massimo Mangino[47,48], Jonathan Marten[38], Evelin Mihailov[109], Michael B. Miller[5], Peter J. van der Most[49], Christopher Oldmeadow[50,51], Antony Payton[52,53], Natalia Pervjakova[54,109], Wouter J. Peyrot[55], Yong Qian[56], Olli Raitakari[57], Rico Rueedi[58,59], Erika Salvi[60], Börge Schmidt[61], Katharina E. Schraut[62], Jianxin Shi[63], Albert V. Smith[64,65], Raymond A. Poot[23], Beate St Pourcain[66,67], Alexander Teumer[68], Gudmar Thorleifsson[28], Niek Verweij[69], Dragana Vuckovic[28], Juergen Wellmann[70], Harm-Jan Westra[71,72,7], Jingyun Yang[73,74], Wei Zhao[75], Zhihong Zhu[181], Behrooz Z. Alizadeh[49,76], Najaf Amin[2], Andrew Bakshi[181], Sebastian E. Baumeister[67,77], Ginevra Biino[78], Klaus Bønnelykke[18], Patricia A. Boyle[72,79], Harry Campbell[61], Francesco P. Cappuccio[80], Gail Davies[32,81], Jan-Emmanuel De Neve[82] , Panos Deloukas[83,84], Ilja Demuth[85,86], Jun Ding[56], Peter Eibich[87,88], Lewin Eisele[60], Niina Eklund[53], David M. Evans[66,182,] , Jessica D. Faul[89], Mary F. Feitosa[90], Andreas J. Forstner[91,92], Ilaria Gandin[28], Bjarni Gunnarsson[93], Bjarni V. Halldórsson[28,94], Tamara B. Harris[95], Andrew C. Heath[96], Lynne J. Hocking[97], Elizabeth G. Holliday[50,51], Georg Homuth[98], Michael A. Horan[99], Jouke-Jan Hottenga[17], Philip L. de Jager[100,101,7], Peter K. Joshi[62], Astanand Jugessur[102], Marika A. Kaakinen[103], Mika Kähönen[104,105], Stavroula Kanoni[82], Liisa Keltigangas-Järvinen[39], Lambertus A.L.M. Kiemeney[27], Ivana Kolcic[106], Seppo Koskinen[54], Aldi T. Kraja[90], Martin Kroh[87], Zoltan Kutalik[107,57,59], Antti Latvala[29], Lenore J. Launer[108], Maël P. Lebreton[11,109,] Douglas F. Levinson[110], Paul Lichtenstein[41], Peter Lichtner[116], David C.M. Liewald[32,81], LifeLines Cohort Study[111], Anu Loukola[29], Pamela A. Madden[96], Reedik Mägi[112], Tomi Mäki-Opas[54], Riccardo E. Marioni[32,113,181], Pedro Marques-Vidal[114], Gerardus A. Meddens[115], George McMahon[66], Christa Meisinger[22], Thomas Meitinger[116], Yusplitri Milaneschi[55], Lili Milani[112], Grant W. Montgomery[117], Ronny Myhre[102], Christopher P. Nelson[30,31], Dale R. Nyholt[118,117], William E.R. Ollier[52], Aarno Palotie[119,7,120,121,122,123,] Lavinia Paternoster[66], Nancy L. Pedersen[41], Katja E. Petrovic[34], David J. Porteous[33], Katri Räikkönen[39,42], Susan M. Ring[66], Antonietta Robino[124], Olga Rostapshova[125], Igor Rudan[62], Aldo Rustichini[126], Veikko Salomaa[54], Alan R. Sanders[127,128], Antti-Pekka Sarin[122,129], Helena Schmidt[130,34], Rodney J. Scott[131,51], Blair H. Smith[132], Jennifer A. Smith[75], Jan A. Staessen[133,134], Elisabeth Steinhagen-Thiessen[85], Konstantin Strauch[135,136], Antonio Terracciano[137], Martin D. Tobin[138], Sheila Ulivi[124], Simona Vaccargiu[24], Lydia Quaye[47], Frank J.A. van Rooij[2,139], Cristina Venturin[i47,48], Anna A.E. Vinkhuyzen[181], Uwe Völker[98], Henry Völzke[68], Judith M. Vonk[49], Diego Vozzi[124], Johannes Waage[18,19], Erin B. Ware[75,140], Gonneke Willemsen[17], John R. Attia[50,51] David A. Bennett[73,74], Klaus Berger[69], Lars Bertram[141,142], Hans Bisgaard[18], Dorret I. Boomsma[17], Ingrid B. Borecki[90], Ute Bultmann[143], Christopher F. Chabris[144], Francesco Cucca[145], Daniele Cusi[60,146], Ian J. Deary[32,81], George V. Dedoussis[40], Cornelia M. van Duijn[2], Johan G. Eriksson[147,42], Barbara Franke[148], Lude Franke[153], Paolo Gasparini[28,124,149], Pablo V. Gejman[127,128], Christian Gieger[21], Hans-Jörgen Grabe[150,151], Jacob Gratten[181], Patrick J.F. Groenen[152], Vilmundur Gudnason[9,65], Pim van der Harst[69,153,154], Caroline Hayward[38,155], David A. Hinds[26], Wolfgang Hoffmann[68], Elina Hypponen[156,157,158],

William G. Iacono[5], Bo Jacobsson[20,102], Marjo-Riitta Järvelin[159,160,161,162], Karl-Heinz Jöckel[61], Jaakko Kaprio[29,122,54], Sharon L.R. Kardia[75], Terho Lehtimäki[163,164], Steven F. Lehrer[165,166], Patrik K.E. Magnusson[41], Nicholas G. Martin[167], Matt McGue[5], Andres Metspalu[112,168], Neil Pendleton[169,170], Brenda W.J.H. Penninx[55], Markus Perola[54,112], Nicola Pirastu[28], Mario Pirastu[24], Ozren Polasek[171,62], Danielle Posthuma[43,172], Christine Power[158], Michael A. Province[90], Nilesh J. Samani[30,31], David Schlessinger[56], Reinhold Schmidt[34], Thorkild I.A. Sørensen[173,15,66], Tim D. Spector[47], Kari Stefansson[93,65], Unnur Thorsteinsdottir[93,65], A. Roy Thurik[1,174,175,3], Nicholas J. Timpson[66], Henning Tiemeier[2,176,177], Joyce Y. Tung[26], André G. Uitterlinden[2], Veronique Vitart[38], Peter Vollenweider[114], David R. Weir[88], James F. Wilson[62,38], Alan F. Wright[38], Dalton C. Conley[178,179], Robert F. Krueger[5], George Davey Smith[66], Albert Hofman[2], David I. Laibson[8], Sarah E. Medland[45], Michelle N. Meyer[180], Jian Yang[181,182], Magnus Johannesson[183], Peter M. Visscher[181,182*], Tõnu Esko[112,6,184,7,#], Philipp D. Koellinger[43,11,3,#], David Cesarini[14,185,#], Daniel J. Benjamin[186,187,#]

[*] These authors contributed equally and are listed in alphabetic order.
[#] Designed and oversaw the study.

**Figure 1 | Manhattan plot for *EduYears* associations (*N* = 293,723).**
The *x*-axis is chromosomal position, and the *y*-axis is the significance on a $-\log_{10}$ scale. The black line shows the genome-wide significance level ($5\times10^{-8}$). The red x's are the approximately 74 independent genome-wide significant associations ("lead SNPs"). The black dots labeled with rs numbers are the 3 Rietveld et al.[7] SNPs.

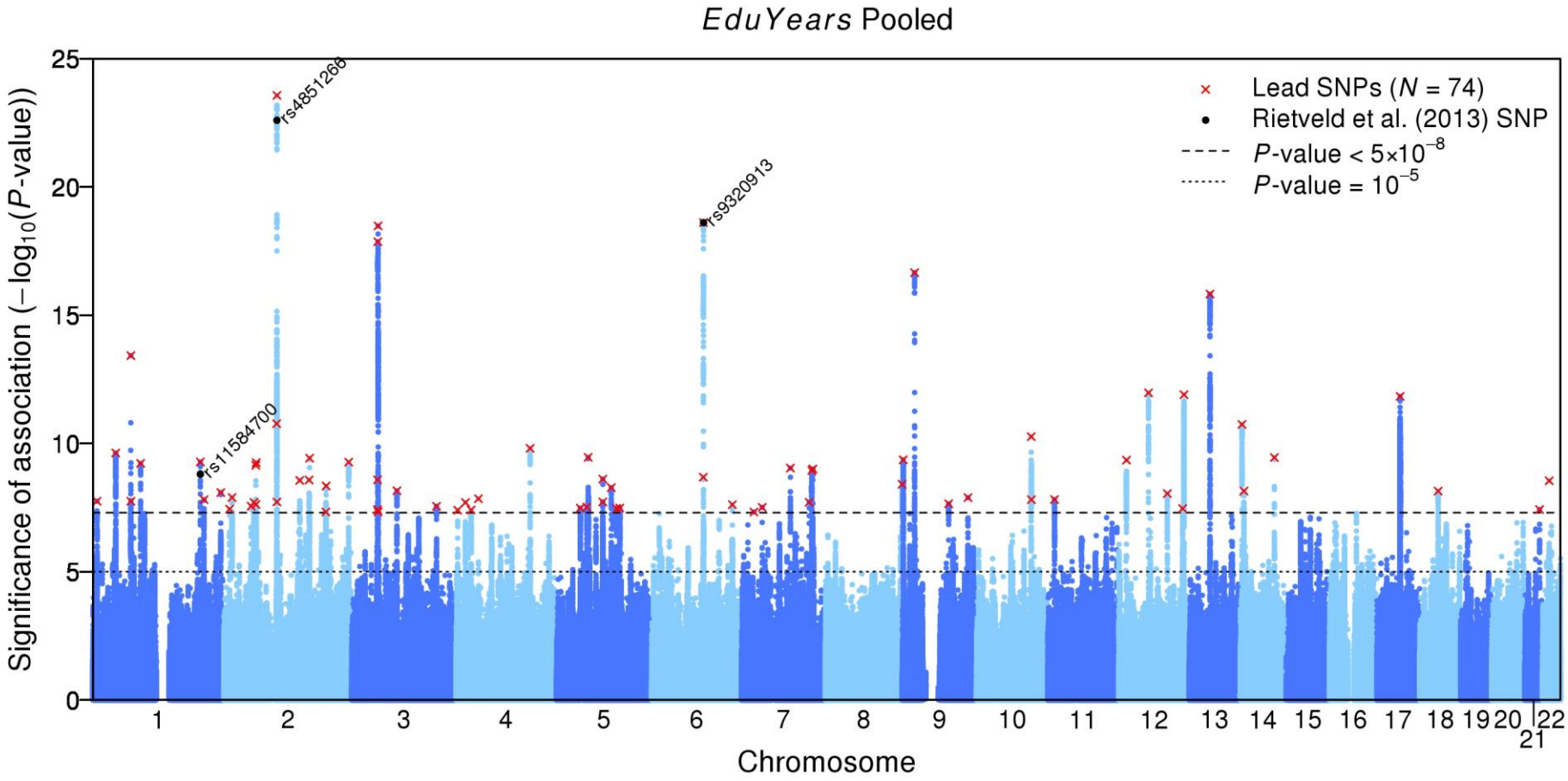

**Figure 2 | Genetic overlap between *EduYears* and other traits.**
**a,** Results from Linkage-Disequilibrium (LD) Score regressions: estimates of genetic correlation with brain volume, neuropsychiatric, behavioral, and anthropometric phenotypes for which GWAS summary statistics were available in the public domain. **b,** Q-Q plots for the 74 lead *EduYears* SNPs in selected phenotypes (for other phenotypes, see Extended Data Fig. 5). SNPs with concordant effects on both phenotypes are pink, and SNPs with discordant effects are blue. SNPs outside the gray area pass Bonferroni-corrected significance thresholds that correct for the total number of SNPs we tested ($P < 0.05/74 = 6.8\times10^{-4}$) and are labeled with their rs numbers. For the sign concordance test: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

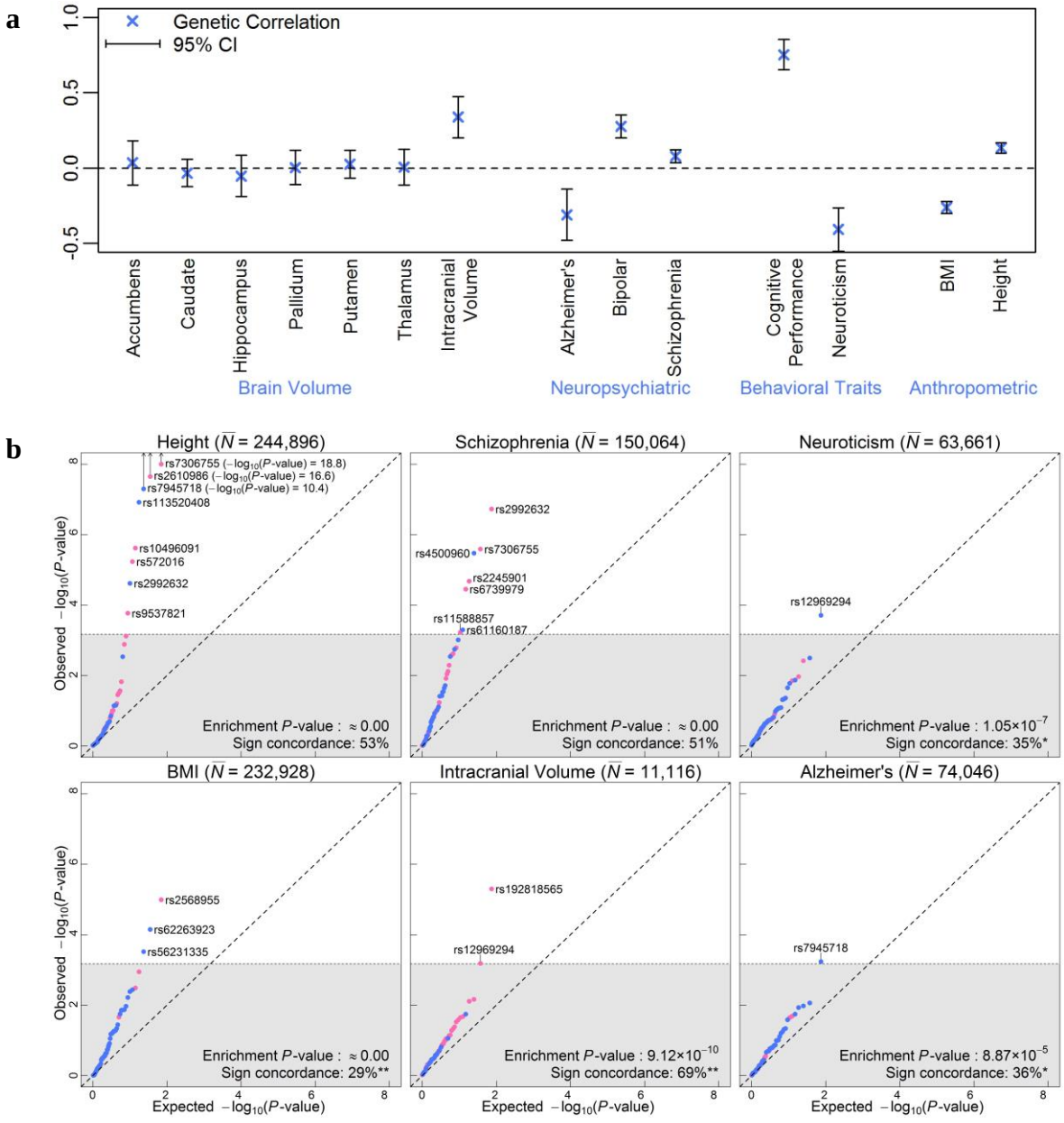

**Figure 3 | Overview of biological annotation**.
**a,** We identified 283 significantly (FDR < 0.05) enriched reconstituted gene sets, which we grouped into the 34 clusters displayed in the graph using the Affinity Propagation method. Each cluster is named after one of its member gene sets. The color represents the *P*-value of the member set exhibiting the most statistically significant enrichment. Overlap between the named representatives of two clusters is represented by an edge. Edge width represents the Pearson correlation $\rho$ between the two respective vectors of gene membership scores ($\rho < 0.3$, no edge; $0.3 \leq \rho < 0.5$, thin edge; $0.5 \leq \rho < 0.7$, intermediate edge; $\rho \geq 0.7$, thick edge). **b,** For each of the plotted disorders, we calculated the odds ratio corresponding to the overlap between genes where *de novo* mutations have been implicated in the disorder and genes nominated by DEPICT as affecting the given GWAS phenotype. The absence of a bar indicates no overlap. *EduYears*, years of education; BMI, body mass index; WHR, waist-to-hip ratio adjusted for BMI. **c,** The genes in the intersection of DEPICT prioritization for *EduYears* and BrainSpan Developmental Transcriptome data (red curve) exhibit higher expression in the brain during prenatal relative to postnatal stages. The prioritized genes exhibit similar median gene-expression levels across different brain areas (gray lines). Analyses were based on median $\log_2$-transformed RNA-Seq data. Error bars represent 95% confidence intervals.

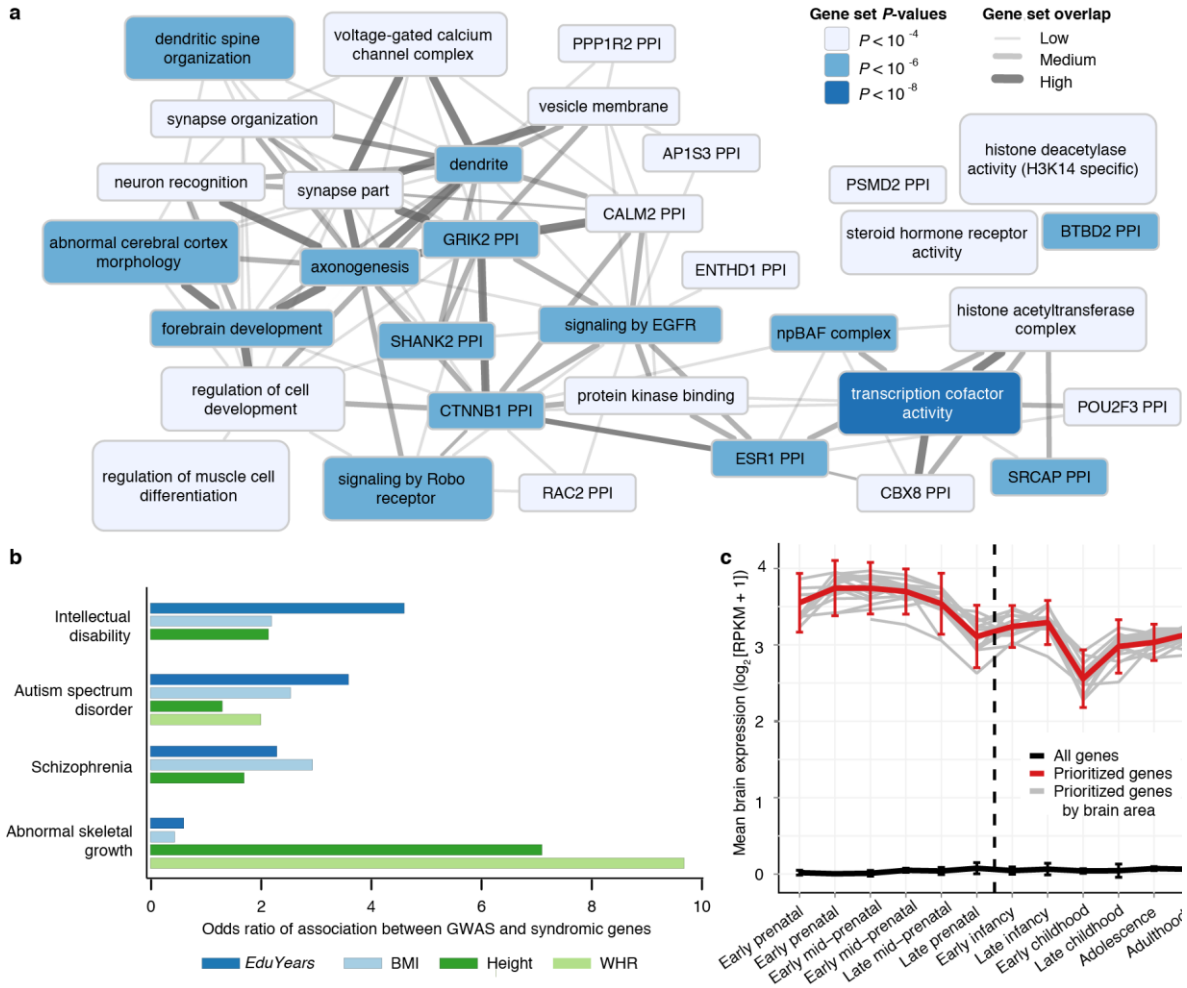

**Figure 4 | Variance explained by *EduYears*-associated SNPs at different levels of significance.** SNPs were selected from the GCTA-COJO analysis with the validation cohort excluded from the meta-analysis. **a**, Predictive power of the polygenic score in unrelated individuals. Cohorts are HRS (pink) and STR (green). The *y*-axis is the variance explained (incremental $R^2$ from including the score in the regression). The line is the sample-size-weighted mean $R^2$. The score is residualized on the first 10 principal components. **b,c**, Partitioning the variance of the SNP-derived polygenic score using a within-family analysis in STR. The polygenic score is not adjusted in **b**, and residualized on the first 10 principal components in **c**. The four variance and covariance components $V_g$, $V_e$, $C_g$ and $C_e$ are defined in Supplementary Information section 3.5.

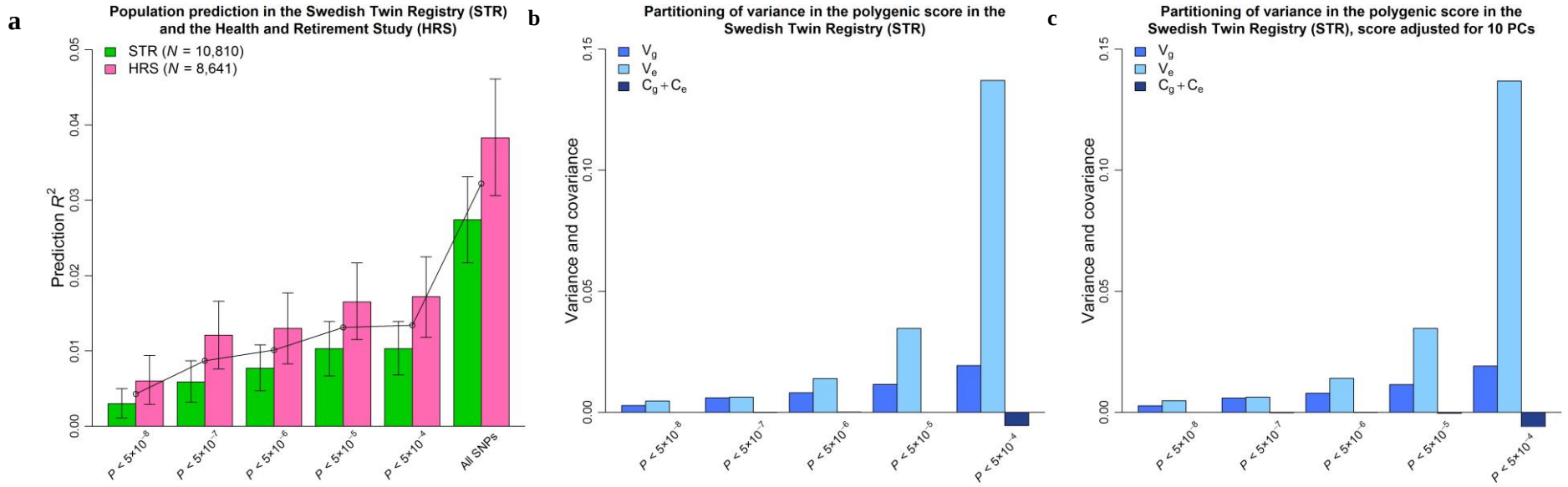

**Table 1 | Significantly prioritized *EduYears*-associated genes.** For each gene we report any of three neuropsychiatric diseases reported to be caused by de novo mutations. We also report up to three of the most significantly enriched non-PPI reconstituted gene sets (other than the MP category decreased fear-related response). When there is more than one independent significant SNP in the DEPICT-defined locus, we give only the one with the lowest GWAS *P*-value. Not all SNPS in the top part of the table reached $P < 5\times10^{-8}$. ID, intellectual disability; ASD, autism spectrum disorder; SCZ, schizophrenia; MP, Mouse Phenome; PPI, InWeb protein-protein interaction subnetwork.

| GWAS locus | Gene symbol | Syndromic gene | Prioritization *P*-value | Lines of evidence | Top-ranking reconstituted gene sets |
|---|---|---|---|---|---|
| **Top 10 genes by DEPICT prioritization (see Supplementary Table 4.1)** | | | | | |
| rs34328009 | *CAMTA1* | – | $4.7\times10^{-13}$ | 2 | abnormal dorsal root ganglion morphology, G alpha (12/13) signaling events |
| rs58852793 | *PPP6R2* | – | $5.2\times10^{-12}$ | 3 | abnormal striatum morphology, histone deacetylase activity (H3K9 specific), NAD-dependent histone deacetylase activity |
| rs35436312 | *FOXO6* | – | $1.2\times10^{-11}$ | 3 | dendrite development, regulation of cell development |
| rs72771875 | *MEF2C* | ID, ASD | $1.4\times10^{-8}$ | 5 | ErbB signaling pathway, abnormal sternum ossificiation, regulation of muscle cell differentiation |
| rs9544418 | *NBEA* | SCZ | $1.8\times10^{-8}$ | 4 | developmental biology, signaling by Robo receptor, dendritic shaft |
| rs12754946 | *USP33* | – | $6.1\times10^{-8}$ | 3 | protein serine/threonine phosphatase complex |
| rs111730030 | *NACC1* | – | $6.8\times10^{-8}$ | 2 | transcription factor binding, kinase binding, regulation of cell development |
| rs11191193 | *MGEA5* | – | $1.5\times10^{-7}$ | 2 | protein kinase binding, kinase binding |
| rs12987662 | *LONRF2* | – | $4.4\times10^{-7}$ | 3 | ectopic cerebellar granule cells, dendritic spine organization, synapse organization |
| rs11712056 | *IP6K2* | – | $6.0\times10^{-7}$ | 2 | transcription coactivator activity, SWI/SNF-type complex, nBAF complex |
| **Top 15 genes in the 74 *EduYears* loci ($P < 5\times10^{-8}$) by lines of evidence (see Supplementary Table 4.5.1)** | | | | | |
| rs4500960 | *TBR1* | ID, ASD | $6.3\times10^{-4}$ | 6 | developmental biology, decreased brain size, abnormal cerebral cortex morphology |
| rs61160187 | *ZSWIM6* | – | $1.3\times10^{-4}$ | 5 | transcription factor binding, negative regulation of signal transduction, PI3K events in ErbB4 signaling |
| rs2457660 | *BCL11A* | ASD | $7.6\times10^{-4}$ | 5 | dendritic spine organization, abnormal hippocampal mossy fiber morphology, SWI/SNF-type complex |
| rs11712056 | *CELSR3* | SCZ | $8.9\times10^{-4}$ | 5 | dendrite morphogenesis, dendrite development, abnormal hippocampal mossy fiber morphology |
| rs192818565 | *MAPT* | ID | $9.5\times10^{-4}$ | 5 | dendrite morphogenesis, abnormal hippocampal mossy fiber morphology, abnormal axon guidance |
| rs7306755 | *SBNO1* | SCZ | $7.2\times10^{-3}$ | 5 | protein serine/threonine phosphatase complex |
| rs12987662 | *NBAS* | – | 0.05 | 5 | – |
| rs1871109 | *SMARCA2* | ID | $8.8\times10^{-6}$ | 4 | – |
| rs11712056 | *MAP4* | ASD | $2.6\times10^{-5}$ | 4 | developmental biology, signaling by Robo receptor, SWI/SNF-type complex |

| | | | | | |
|---|---|---|---|---|---|
| rs10061788 | *LINC00461* | – | $6.8\times10^{-5}$ | 4 | decreased brain size, abnormal cerebral cortex morphology, abnormal hippocampal mossy fiber morphology |
| rs9320913 | *POU3F2* | – | $3.6\times10^{-4}$ | 4 | dendrite morphogenesis, developmental biology, decreased brain size |
| rs11712056 | *RAD54L2* | SCZ | $4.8\times10^{-4}$ | 4 | decreased brain size, SWI/SNF-type complex, nBAF complex |
| rs2964197 | *PLK2* | – | $8.1\times10^{-4}$ | 4 | negative regulation of signal transduction, PI3K events in ErbB4 signaling |
| rs9537821 | *PCDH17* | – | $1.6\times10^{-3}$ | 4 | activation of RAC, abnormal axon guidance, axon guidance (Reactome) |

[1] Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA, Rotterdam, The Netherlands
[2] Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
[3] Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam 3062 PA, The Netherlands
[4] Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA
[5] Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA
[6] Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA 2116, USA
[7] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[8] Department of Economics, Harvard University, Cambridge, MA 02138, USA
[9] Icelandic Heart Association, Kopavogur, 201, Iceland
[10] Faculty of Pharmaceutical Sciences, University of Iceland, 107 Reykjavík, Iceland
[11] Amsterdam Business School, University of Amsterdam, Amsterdam, 1018 TV, The Netherlands
[12] Department of Government, Uppsala University, Uppsala, 751 20, Sweden
[13] New York Genome Center, New York, NY 10013, USA
[14] Department of Economics, New York University, New York, NY 10012, USA
[15] The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Faculty of Health and Medical Sciences, København, 2100, Denmark
[16] Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark Lyngby, 2800, Denmark
[17] Department of Biological Psychology, VU University Amsterdam, Amsterdam, 1081 BT, The Netherlands
[18] Copenhagen Prospective Studies on Asthma in Childhood, Faculty of Health and Medical Sciences, University of Copenhagen, Gentofte, 2820, Denmark
[19] Steno Diabetes Center, 2820 Gentofte, Denmark
[20] Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg, SE 416 85, Sweden
[21] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany
[22] Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany
[23] Department of Cell Biology, Erasmus Medical Center Rotterdam, 3015 CN, The Netherlands
[24] Istituto di Ricerca Genetica e Biomedica U.O.S. di Sassari, National Research Council of Italy, Sassari, 07100, Italy
[25] Psychology, University of Illinois, IL 61820, Champaign, USA
[26] 23andMe, Inc., Mountain View, CA 94041, USA
[27] Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands
[28] Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, 34100, Italy
[29] Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland
[30] Department of Cardiovascular Sciences, University of Leicester, Leicester, LE3 9QP, UK
[31] NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP, UK
[32] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
[33] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
[34] Department of Neurology, General Hospital and Medical University Graz, Graz, 8036, Austria
[35] Institute for Medical Informatics, Statistics and Documentation, General Hospital and Medical University Graz, Graz, 8036, Austria
[36] Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, OX3 7LE, UK
[37] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
[38] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
[39] Institute of Behavioural Sciences, University of Helsinki, Helsinki, FI-00014, Finland
[40] Nutrition and Dietetics, Health Science and Education, Harokopio University, Athens, 17671, Greece
[41] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 171 77, Sweden
[42] Folkhälsan Research Centre, Helsingfors, FI-00014, Finland

[43] Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, The Netherlands
[44] Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, 6525 EC, The Netherlands
[45] Quantitative Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia
[46] Lifespan Psychology, Max Planck Institute for Human Development, Berlin, 14195, Germany
[47] Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK
[48] NIHR Biomedical Research Centre, Guy's and St. Thomas' Foundation Trust, London, SE1 7EH, UK
[49] Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, The Netherlands
[50] Public Health Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia
[51] Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW 2300, Australia
[52] Centre for Integrated Genomic Medical Research, Institute of Population Health, The University of Manchester, Manchester, M13 9PT, UK
[53] School of Psychological Sciences, The University of Manchester, Manchester, M13 9PL, UK
[54] Department of Health, THL-National Institute for Health and Welfare, Helsinki, FI-00271, Finland
[55] Psychiatry, VU University Medical Center & GGZ inGeest, Amsterdam, 1081 HL, The Netherlands
[56] Laboratory of Genetics, National Institute on Aging, Baltimore, MD 21224, USA
[57] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20521, Finland
[58] Department of Medical Genetics, University of Lausanne, Lausanne, 1005, Switzerland
[59] Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
[60] Department Of Health Sciences, University of Milan, Milano, 20142, Italy
[61] Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, 45147, Germany
[62] Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK
[63] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892-9780, USA
[64] Icelandic Heart Association, Kopavogur, 201, Iceland
[65] Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
[66] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK
[67] School of Oral and Dental Sciences, University of Bristol, Bristol, BS1 2LY, UK
[68] Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany
[69] Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, 9700 RB, The Netherlands
[70] Institute of Epidemiology and Social Medicine, University of Muenster, Muenster, 48149, Germany
[71] Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, MA 02115, Boston, USA
[72] Partners Center for Personalized Genetic Medicine, Boston, MA 02115, USA
[73] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA
[74] Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612, USA
[75] Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA
[76] Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, 9713 GZ, The Netherlands
[77] Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, D-93053, Germany
[78] Institute of Molecular Genetics, National Research Council of Italy, Pavia, 27100, Italy
[79] Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL 60612, USA
[80] Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK
[81] Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
[82] Saïd Business School, University of Oxford, Oxford, OX1 1HP, UK
[83] William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK
[84] Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[85] The Berlin Aging Study II; Research Group on Geriatrics, Charité – Universitätsmedizin Berlin, Germany, Berlin, 13347, Germany
[86] Institute of Medical and Human Genetics, Charité-Universitätsmedizin, Berlin, Berlin, 13353, Germany
[87] German Socio- Economic Panel Study, DIW Berlin, Berlin, 10117, Germany

[88] Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK
[89] Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48109, USA
[90] Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO 63018, USA
[91] Institute of Human Genetics, University of Bonn, Bonn, 53127, Germany
[92] Department of Genomics, Life and Brain Center, University of Bonn, Bonn, 53127, Germany
[93] deCODE Genetics/Amgen Inc., Reykjavik, IS-101, Iceland
[94] Institute of Biomedical and Neural Engineering, School of Science and Engineering, Reykjavik University, Reykjavik 101, Iceland
[95] Laboratory of Epidemiology, Demography, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, United States
[96] Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA
[97] Division of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, UK
[98] Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, 17475, Germany
[99] Manchester Medical School, The University of Manchester, Manchester, 9PT, UK
[100] Program in Translational NeuroPsychiatric Genomics, Departments of Neurology & Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA
[101] Harvard Medical School, Boston, MA 02115, USA
[102] Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, N-0403, Norway
[103] Department of Genomics of Common Disease, Imperial College London, London, W12 0NN, UK
[104] Department of Clinical Physiology, Tampere University Hospital, Tampere, 33521, Finland
[105] Department of Clinical Physiology, University of Tampere, School of Medicine, Tampere, 33014, Finland
[106] Public Health, Medical School, University of Split, 21000 Split, Croatia
[107] Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, 1010, Switzerland
[108] Neuroepidemiology Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, USA
[109] Amsterdam Brain and Cognition Center, University of Amsterdam, 1018 XA, Amsterdam, The Netherlands
[110] Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305-5797, USA
[111] LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, Groningen, 9713 BZ, The Netherlands
[112] Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia
[113] Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
[114] Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, 1011, Switzerland
[115] Tema BV, 2131 HE Hoofddorp, The Netherlands
[116] Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany
[117] Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia
[118] Institute of Health and Biomedical Innovation, Queensland Institute of Technology, Brisbane, QLD 4059, Australia
[119] Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[120] The Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
[121] Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA
[122] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, 00014, Finland
[123] Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA
[124] Medical Genetics, Institute for Maternal and Child Health IRCCS "Burlo Garofolo", Trieste, 34100, Italy
[125] Harvard Kennedy School, Harvard University, Cambridge, MA 02139
[126] Department of Economics, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA
[127] Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL 60201-3137, USA
[128] Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA
[129] Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki 00300, Finland

[130] Research Unit for Genetic Epidemiology, Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, General Hospital and Medical University, Graz, Graz, 8010, Austria
[131] Information Based Medicine Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia
[132] Medical Research Institute, University of Dundee, Dundee, DD1 9SY, UK
[133] Research Unit Hypertension and Cardiovascular Epidemiology, Department of Cardiovascular Science, University of Leuven, Leuven, 3000, Belgium
[134] R&D VitaK Group, Maastricht University, Maastricht, 6229 EV, The Netherlands
[135] Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany
[136] Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians-Universität, Munich, 81377, Germany
[137] Department of Geriatrics, Florida State University College of Medicine, Tallahassee, FL 32306, USA
[138] Department of Health Sciences and Genetics, University of Leicester, Leicester, LE1 7RH, UK
[139] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
[140] Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA
[141] Platform for Genome Analytics, Institutes of Neurogenetics & Integrative and Experimental Genomics, University of Lübeck, Lübeck, 23562, Germany
[142] Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK
[143] Department of Health Sciences, Community & Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, 9713 AV, The Netherlands
[144] Department of Psychology, Union College, Schenectady, NY 12308, USA
[145] Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, 9042, Italy
[146] Institute of Biomedical Technologies, Italian National Research Council, Segrate (Milano), 20090, Italy
[147] Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, 00014, Finland
[148] Departments of Human Genetics and Psychiatry, Donders Centre for Neuroscience, Nijmegen, 6500 HB, The Netherlands
[149] Sidra, Experimental Genetics Division, Sidra, Doha 26999, Qatar
[150] Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, 17475, Germany
[151] Department of Psychiatry and Psychotherapy, HELIOS-Hospital Stralsund, Stralsund, 18437, Germany
[152] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, 3062 PA, The Netherlands
[153] Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen. 9700 RB, The Netherlands
[154] Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, 1105 AZ, The Netherlands
[155] Generation Scotland, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
[156] School of Population Health and Sansom Institute, University of South Australia, SA5000, Adelaide, Australia
[157] South Australian Health and Medical Research Institute, Adelaide, SA5000, Australia
[158] Population, Policy and Practice, UCL Institute of Child Health, London, WC1N 1EH, UK
[159] Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, W2 1PG, UK
[160] Center for Life Course Epidemiology, Faculty of Medicine, University of Oulu, Oulu, FI-90014 , Finland
[161] Unit of Primary Care, Oulu University Hospital, Oulu, 90029 OYS, Finland
[162] Biocenter Oulu, University of Oulu, FI-90014 Oulu, Finland
[163] Fimlab Laboratories, Tampere, 33520, Finland
[164] Department of Clinical Chemistry, University of Tampere, School of Medicine, Tampere, 33014, Finland
[165] Economics, NYU Shanghai, 200122, Pudong, China
[166] Policy Studies, Queen's University, Kingston, K7L3N6, Canada
[167] Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia
[168] Institute of Molecular and Cell Biology, University of Tartu, Tartu, 51010, Estonia
[169] Centre for Clinical and Cognitive Neuroscience, Institute Brain Behaviour and Mental Health, Salford Royal Hospital, Manchester, M6 8HD, UK
[170] Manchester Institute Collaborative Research in Ageing, University of Manchester, Manchester, M13 9PL, UK

[171] Faculty of Medicine, University of Split, Croatia, Split 21000, Croatia
[172] Department of Clinical Genetics, VU Medical Centre, Amsterdam, 1081 HV, The Netherlands
[173] Institute of Preventive Medicine, Frederiksberg Hospital, The Capital Region, Frederiksberg, 2000, Denmark
[174] Montpellier Business School, Montpellier, 34080, France
[175] Panteia, Zoetermeer, 2715 CA, The Netherlands
[176] Department of Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
[177] Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
[178] Department of Sociology, New York University, New York, NY 10012, USA
[179] School of Medicine, New York University, NY 10016, New York, USA
[180] Bioethics Program, Union Graduate College - Icahn School of Medicine at Mount Sinai, Schenectady, NY 12308, USA
[181] Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia
[182] The University of Queensland Diamantina Institute, The Translation Research Institute, Brisbane, QLD 4102, Australia
[183] Department of Economics, Stockholm School of Economics, Stockholm, 113 83, Sweden
[184] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[185] Research Institute for Industrial Economics, Stockholm, 10215, Sweden
[186] Department of Economics, Cornell University, Ithaca, NY 14853, USA
[187] Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089-3332, USA